Engineered reciprocal chromosome translocations drive high threshold, reversible population replacement in Drosophila

Anna B. Buchman*#, Tobin Ivy*, John M. Marshall$, Omar S. Akbari*#&, and Bruce A. Hay*&

*Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA. $School of Public Health, University of California Berkeley, Berkeley CA. #Department of Entomology, University of California, Riverside, Riverside, CA.

&Authors for correspondence
haybruce@caltech.edu
omar.akbari@ucr.edu

Abstract:

Replacement of wild insect populations with transgene-bearing individuals unable to transmit disease or survive under specific environmental conditions provides self-perpetuating methods of disease prevention and population suppression, respectively. Gene drive mechanisms that require the gene drive element and linked cargo exceed a high threshold frequency to spread are attractive because they offer several points of control: they bring about local, but not global population replacement; and transgenes can be eliminated by reintroducing wildtypes into the population so as to drive the frequency of transgenes below the threshold required for drive. It has long been recognized that reciprocal chromosome translocations could, in principal, be used to bring about high threshold gene drive through a form of underdominance. However, translocations able to drive population replacement have not been reported, leaving it unclear if translocation-bearing strains fit enough to mediate gene drive can easily be generated. Here we use modeling to identify a range of conditions under which translocations should spread, and the equilibrium frequencies achieved, given specific introduction frequencies, fitness costs and migration rates. We also report the creation of engineered translocation-bearing strains of *Drosophila melanogaster*, generated through targeted chromosomal breakage and homologous recombination. By several measures translocation-bearing strains are fit, and drive high threshold, reversible population replacement in laboratory populations. These observations, together with the generality of the tools used to generate translocations, suggest that engineered translocations may be useful for controlled population replacement in many species.

Insects act as vectors for a number of important diseases of humans, animals, and plants (1). Traditional vector control is often challenging, with the degree of protection provided being proportional to the effort put into control. In addition, depending on the environment, specific methods of vector control, such as environment modification or use of insecticides, may be impractical or have undesirable side effects. A complementary strategy for disease prevention, first articulated many decades ago (2), involves using gene drive to bring about replacement of wild, disease transmitting insect populations with individuals engineered to be refractory to disease transmission, but still subject to traditional vector control (reviewed in (3-6). In a variant of this idea, population replacement has also been proposed as a method for bringing about disease prevention and/or a reduction in insect mediated damage through periodic population suppression (7, 8). This can occur when replacement results in all individuals carrying genes that cause death or failure to diapause in response to application of an otherwise benign chemical, or a seasonal change in an environmental variable such as temperature or humidity. An important appeal of these strategies is that they are species-specific and potentially self-perpetuating.

Because transgenes that mediate disease resistance or conditional lethality are unlikely to confer a fitness benefit to carriers an essential component of most population replacement strategies (see (9-12) for several non-drive based replacement strategies) is linkage with a gene drive mechanism that carries transgenes to high frequency following release. These drive mechanisms must be strong enough to spread genes to high frequency in wild populations on human timescales, while also functioning within regulatory frameworks (13-17). Central to the latter are issues of confinement and reversibility: can the spread of transgenes to high frequency be limited to locations in which their presence is sought, and can the population be restored to the pre-transgenic state?

An important characteristic of any gene drive mechanism that relates to the above questions is its level of invasiveness: its ability to increase in frequency

both at the release site and in surrounding areas linked to the release site by various levels of migration, when introduced at various population frequencies. Low threshold gene drive mechanisms require that only a small fraction of individuals in the population carry the drive element in order for spread to occur locally (18, 19). Examples include engineered *Medea* chromosomal elements (20-22), several other possible single locus chromosomal elements (23), site-specific nucleases that home into their target site (24-29), and site-specific nucleases that result in sex ratio distortion (30). These mechanisms are predicted to be invasive because low levels of migration of drive element-bearing individuals into areas outside the release area may, depending on the threshold and the migration rate (18, 19, 31), result in these areas being seeded with enough transgene-bearing individuals for drive to occur. Low threshold, invasive gene drive mechanisms are attractive when the goal is to spread transgenes over a large area, and migration rates between the release site and surrounding areas of interest are low. However, for these same reasons, it is likely to be challenging to restore the population to the pre-transgenic state if desired. Given the intense scrutiny with which releases of insects engineered to suppress population numbers while ultimately disappearing from the population have been greeted (15-17), gene drive mechanisms that have a limited capacity to spread, and that can easily be eliminated from the population, thereby restoring the population to a pre-transgenic state, may be useful in some contexts.

High (or higher) threshold gene drive mechanisms require, as their name implies, that transgenes make up a much larger fraction of the total insect population (important examples range from 15-70%) before gene drive occurs. Below this frequency transgenes are instead actively eliminated from the population. In short, these drive mechanisms behave as a frequency-dependent bistable switch. High transgene frequencies are needed to initiate drive at the release site, limiting the possibility that unintended release of a few individuals could initiate replacement. Once replacement has occurred at the release site, spread to high frequency in areas connected to the release site by low levels of migration is prevented because the transgene never reaches the threshold

105 frequency needed for drive. Finally, transgenes can be eliminated from the
106 population if the release of wildtypes results in the frequency of transgenics being
107 driven below the threshold required for drive.

108 A number of gene drive mechanisms that could in principal bring about local and
109 reversible population replacement have been proposed. Examples include a
110 number of single locus gene drive mechanisms (23, 32, 33), reciprocal
111 chromosome translocations, inversions and compound chromosomes (34), and
112 several forms of engineered underdominance (23, 35-39) (40). One of these,
113 UD^{MEL} (double *Medea*), has recently been shown to drive reversible population
114 replacement into populations of wildtype Drosophila (38). A second system has
115 been shown to drive high threshold population replacement in Drosophila in a
116 split configuration (40). In each of these systems gene drive occurs when
117 transgene-bearing chromosomes experience frequency-dependent changes in
118 fitness with respect to non-transgene-bearing counterparts, with the former
119 having high fitness at high frequency and lower fitness at low frequency. These
120 systems all rely, in one way or another, on the phenomena of underdominance,
121 in which transgene-bearing heterozygotes (or some fraction of them or their
122 progeny) have a lower fitness than either homozygous wildtypes or homozygous
123 transgenics (or transgene-bearing trans-heterozygote in some three allele
124 cases). If the frequency of one allele or pair of alleles or chromosome type is
125 above a critical threshold it spreads to genotype, and in some cases allele
126 fixation. Conversely, if it falls below the critical threshold it is lost in favor of the
127 other allele or chromosome type, usually wildtype. In broad outline, this behavior
128 occurs because when transgene-bearing individuals are common they mate
129 mostly with each other, producing transgene-bearing offspring of high fitness
130 (high survival and/or fecundity), while wildtypes mate mostly with transgene-
131 bearing individuals, producing a preponderance of heterozygous offspring of low
132 fitness (inviable and/or with reduced fecundity). However, when the frequency of
133 wildtypes is high the tables are turned, with transgene-bearing individuals
134 producing high frequencies of unfit heterozygous progeny, and wildtypes
135 producing a high frequency of fit homozygous progeny.

Here we focus on the use of engineered reciprocal chromosome translocations as a high threshold gene drive mechanism. Reciprocal chromosome translocations were the first gene drive mechanism proposed (2). Their structure and genetic behavior are illustrated in Figure 1A. A reciprocal chromosome translocation results in the mutual exchange of DNA between two non-homologous chromosomes (41). Provided that the translocation breakpoints do not alter the expression and/or function of nearby genes, translocation heterozygotes and homozygotes can in principal be phenotypically normal. Thus, phenotypically normal, naturally occurring translocation-bearing individuals are found in populations of many species (42), including humans (43, 44). However, translocation heterozygotes are usually semisterile, producing a high frequency of inviable offspring. This occurs because meiosis in a translocation heterozygote can generate a variety of different products. Three patterns of segregation are possible: alternate, adjacent-1 and adjacent-2 (Figure 1A). While alternate segregation leads to the production of gametes with a full genome complement, adjacent-1 and adjacent-2 segregation lead to the production of aneuploid gametes, resulting in the death of progeny that inherit an unbalanced chromosome set. In many species alternate and adjacent-1 segregation occur roughly equally, with adjacent-2 segregation being rare (45, 46). In such species progeny genotypes and survival phenotypes resulting from crosses between translocation-bearing individuals and wildtypes are as illustrated in the Punnett square in Figure 1B. Progeny with unbalanced genotypes die, while balanced translocation heterozygotes, translocation homozygotes, and homozygous wildtypes survive.

In 1940 Serebrovski proposed that the release of homozygous translocation-bearing males could be used to drive population suppression because many progeny would be semisterile, thereby driving down population fitness over multiple generations (47). He, Dobzhansky (48), and later Curtis (2), also noted that the frequency of translocations lacks a stable internal equilibrium, with either wildtype or translocation-bearing chromosomes spreading to fixation in an isolated population through natural selection (differential survival of the relevant

chromosome type) if their frequency rose above 50%, for a translocation with no fitness cost to carriers. Curtis proposed that if a gene beneficial to humans could be linked to the translocation breakpoint, this behavior of translocations could be used to spread the gene into the wild population. Whitten subsequently noted that the same approach could be used to spread a trait conferring conditional lethality, which could be used to bring about population suppression (7). More recent modeling work has highlighted the potential of translocations for bringing about local, but not global population replacement, and the ease of reversal (19).

Though it is clear from evolutionary studies that translocations can become fixed in populations (42), efforts to directly bring about population replacement using translocations created in the lab have not been successful (34, 49-51). There may be several reasons for this. First, translocation-bearing individuals (particularly homozygotes) generated in the past typically had very low fitness, probably at least in part because they were generated using X-rays, which can result in a high frequency of background mutations. Second, more recently it has become clear that chromosome positioning and structure in the nucleus can play a role in determining large-scale patterns of gene expression, and that chromosome translocation can result in changes in the patterns of gene expression (52, 53). These latter observations leave it fundamentally unclear whether translocation-bearing individuals of high fitness can be easily generated, even if the breakpoints involved are located in gene deserts. For example, it could be that phenotypically normal translocation-bearing individuals observed in nature simply represent the relatively rare cases in which chromosome rearrangement does not result in fitness being compromised. To explore these issues, and to determine if translocation-based gene drive can be used to bring about population replacement, we first use modeling to explore the relationship between variables such as introduction frequency, fitness cost, and reciprocal migration with non-target populations containing widltypes, for the ability of a translocation to spread, and the equilibrium frequencies achieved in replaced and surrounding populations. We then describe a general approach to generation and identification of site-specific reciprocal chromosomal translocations. Finally, we

provide the first demonstration that engineered translocations are capable of bringing about threshold-dependent population replacement, in *Drosophila melanogaster*.

Some predicted characteristics of translocation-based gene drive.

Early modeling work by Serebrovskii and Curtis showed that if a translocation results in no fitness cost to carriers, and is present in a population experiencing no incoming migration of wildtypes, it will spread to allele fixation when present at population frequencies greater than 50%, and will be eliminated when present at lower frequencies (2). Curtis also noted briefly that translocations that resulted in a fitness cost to carriers could still spread to allele fixation, but the threshold introduction frequency would be increased (2). Given the past failures to bring about translocation-mediated population replacement noted above, and the likelihood that chromosome translocation itself and/or the GOI placed at the breakpoints will result in some fitness cost to carriers, we sought to understand more generally how fitness cost affects translocation spread. The time to allele fixation is particularly relevant for contexts in which the goal is to ultimately bring about population suppression in response to a seasonal variable such as temperature or humidity.

In figure 2A we illustrate the relationship between fitness cost, introduction frequency and time to translocation allele fixation (approximated as the point at which >99% of individuals carry at least one translocation copy), for a single introduction into an isolated population. The plot illustrates several important points. First, whenever translocations spread, they spread to fixation relatively quickly, with the time needed being inversely related to the introduction frequency. Second, translocations that confer large fitness costs to carriers can also spread rapidly, so long as the introduction frequency is increased. The plot in Figure 2B illustrates a related case in which the translocation is introduced over three generations at the specified frequency. It shows that with modest extra effort rapid drive can be achieved, even for very high fitness costs. While these

introduction frequencies represent a large percentage of the wild population, they are still much lower than those used in self-limiting genetic population suppression strategies such as SIT and RIDL (54), and unlike SIT and RIDL, result in sustained changes to the population.

In real world scenarios other than initial field-testing - in which population isolation will be essential - there is likely to be some level of reciprocal migration between the target area (source population 1) and surrounding areas (population 2) containing wildtypes. Marshall and Hay showed that for realistic population sizes (>1000 individuals), there are no reciprocal migration rates that support population replacement in a second, wildtype-containing population (population 2) linked to a source population (population 1) in which replacement is initiated. Due to the high frequency of death among the progeny of translocation-bearing individuals that mate with wildtype, the frequency of translocation-bearing individuals in population 2 never rises to a level that supports drive (see also Figure 3A, C). Instead, when migration rates are high (~6.8%, or lower when the translocation is associated with a fitness cost), translocations are eliminated from both populations (19). Here we consider a related question: what effect does reciprocal migration have on the characteristics of population replacement in the target population, and the genotypic composition of neighboring populations linked by migration, in which drive does not occur?

We consider a specific scenario in which three populations are linked in series: the target population (population 1) is linked to a second population consisting initially of wildtypes (population 2) through migration; population 2 is also linked through migration to a third population consisting initially of wildtypes (population 3), which is not linked directly with population 1. We ask what the equilibrium frequencies are in each population for different levels of migration? In the case of a low threshold gene drive mechanism such as *Medea* or homing by a HEG, the equilibrium frequency in population 1 will approach fixation since these drive elements spread invasively into surrounding populations connected to the target

population by low levels of migration. In contrast, the situation for high threshold gene drive mechanisms is fundamentally different since wildtypes will, by definition, always be present in surrounding non-target populations in which transgene levels sufficient for drive are not achieved. Previous modeling studies of underdominant systems have noted that the presence of reciprocal migration can result in internal equilibria containing both wldtype and underdominant alleles (36, 37) (55). Here we consider the case of reciprocal translocations specifically.

Figure 3A illustrates a specific scenario, in which a translocation with no fitness cost is introduced into population 1 at a frequency of 70%, and is connected to a similarly sized population 2 by a migration rate of 1%. Population 2 is connected to a similarly sized population 3 by the same migration rate. The translocation spreads to high frequency (99%) in population 1, but not to allele or genotype fixation, since wildtypes are introduced into population 1 each generation. Translocation-bearing genotypes are also present at modest levels (<5% (4.954%) in population 2, and <1% (0.08116%) in population 3. Figure 3A also illustrates an identical scenario in which the migration rate is now 5%. In this case the translocation equilibrium frequency is <95% (94.55%) in population 1, <23% (22.58%) in population 2, and ~2% (2.031%) for population 3. The general relationship between fitness cost, migration rate and equilibrium frequency in population 1 is illustrated in Figure 3B. The highest level of incoming wildtype migration that can be tolerated for a translocation with no fitness cost (~6.8% / generation) results in an equilibrium translocation genotype frequency of ~90% in population 1. Decreased levels of migration result in correspondingly higher equilibrium frequencies, which approach fixation as the migration rate falls to zero (as in Figure 2). Populations 2 (Figure 3C) and 3 (Figure 3D) show the opposite behavior. As migration rate increases, the fraction of translocation-bearing individuals increases in population 2, reaching a maximum of ~25% for a translocation with no fitness cost and migration rate of 6.8%. However, for similar migration rates the fraction of translocation-bearing individuals in population 3 is

dramatically reduced. Increased fitness costs result in a minimal decrease in equilibrium translocation frequency in all three populations (Figure 3B-D).

These observations illustrate a fundamental set of tradeoffs associated with high threshold gene drive. While drive can be spatially limited to a single population, this comes with a cost: the continuous introduction of wildtypes from neighboring populations, which keeps the equilibrium frequency of transgene-bearing individuals below 100%. Depending on the disease system being considered, the presence of some level of non-transgene-bearing individuals within the target area may have important epidemiological consequences, as a residual population of wildtype mosquitoes may be capable of sustaining transmission, although this remains to be investigated. Population suppression following activation of condition-dependent lethality may also be challenging in the face of significant levels of wildtype migration. Finally, the presence of some level of translocation-bearing individuals outside the target area may have regulatory implications even if these levels are insufficient for drive. That said, any such issues are likely to be local since the decrease in frequency of drive element-bearing individuals in underdominant systems drops off rapidly in a series of linked populations (Figure 3B-D). Together, these observations suggest that high threshold gene drive is likely to be most epidemiologically effective and able to satisfy regulatory requirements relating to the presence and movement of transgene-bearing organisms within target areas circumscribed by significant barriers to migration.

Engineering Reciprocal Translocations in Drosophila

Cells or organisms carrying translocations with defined breakpoints have recently been generated using several strategies. One set of approaches begins with two non-homologous chromosomes that each have a different transgene-bearing cassette inserted at a specific position. Recombination between the two chromosomes to generate a translocation is then driven by FLP/FRT recombination (56), Cre/*loxP* recombination (57, 58), or homologous

recombination following double-stranded break creation within the transgene cassettes using a site-specific nuclease (58-60). Translocations have also been generated in completely wildtype backgrounds, following Crispr/Cas9-mediated cleavage of two otherwise wildtype chromosomes followed by non-homologous end joining (61-63). In this latter case, PCR-based methods were used to identify pools of cells or individuals carrying translocations.

We sought to create translocations using a variant of the approach described by Egli et al. in which homologous recombination between two chromosomes follows double-stranded break creation using the rare-cutting site-specific nuclease I-SceI (58). However, rather than use their approach for identification of potential translocation bearing individuals, which involves scoring for the loss of the marker *y*+ in an otherwise a *y*- background, we created a system in which recombination results in the creation of a dominant marker. This approach can be used in otherwise wildtype genetic backgrounds, in diverse species.

Two constructs (A and B) were generated (Figure 4B). Each construct includes several components. These include (from left to right) a transformation marker (the *white* gene); a location that could be used as an insertion point of a gene of interest (GOI); a promoter that drives the expression of a dominant florescent marker, either ubiquitously (the Opie2 viral promoter, (64) or in oenocytes (65); a splice donor site, and two stretches of DNA used as substrates for homologous recombination, annotated as UVW and XYZ, each roughly 670bp in length. These DNA fragments were derived from the mouse IgG locus, and thus lack homology with the Drosophila genome. Two target sites for the rare cutting homing endonuclease I-SceI were inserted between UVW and XYZ. To the right of these elements were positioned a splice acceptor, a promoterless reporter gene (GFP or dsRed), and a phiC31 recombination attB site.

These constructs were introduced into flies at three separate attP locations: construct A at 51C on chromosome 2, and construct B at 68E or 70A2 on

chromosome 3 (Figure 4A). The attP insertion sites at 51C and 68E lie some distance from annotated genes, while the 70A2 site lies within a cluster of tRNA loci. Both constructs were oriented in the same direction with respect to their centromeres (Figure 4A). The constructs were designed so that flies bearing construct A, located on the second chromosome, would express the svp-driven eGFP marker, while construct B, located on the third chromosome, would express the opiap2-driven dsRED marker (Figure 4B). Transgenics for construct B behaved as expected, and were dsRED positive throughout their body. However, transgenics for construct A had no detectable GFP expression. The basis for this is unclear, but could be due to inappropriate splicing of the XYZ-UVW sequence in this construct. Regardless, as illustrated below, one marker is sufficient to identify translocation-bearing individuals.

To generate translocation-bearing individuals we created stocks doubly homozygous for constructs A and B (51C; 71A2 or 51C; 68E). These were then mated with flies that express I-SceI under the control of the Hsp70 heat shock promoter (66). Progeny carrying all three transgenes were subjected to multiple rounds of heat shock during larval stages and as adults. Adults were outcrossed to wildtype, and progeny examined under a fluorescent dissecting scope. In a number of individuals strong ubiquitous GFP expression was observed. This is the predicted outcome if I-SceI expression results in cleavage of both transgene-bearing chromosomes (Fig. 4C), followed by homologous recombination between XYZ- and UVW-bearing ends of the two different chromosomes (Fig. 4D,E). Putative translocation heterozygotes (T_1/+; T_2/+) were individually mated to wild type individuals (+/+; +/+) to generate males and female translocation heterozygotes (identified as GFP-expressing). These were mated with each other to generate putative translocation homozygotes (T_1/ T_1; T_2/ T_2). PCR and sequencing of products from genomic DNA of these individuals was used to demonstrate that these individuals were homozygous for both translocation products (Methods and Figure 4F).

To explore the genetic behavior of translocation-bearing chromosomes and the fitness of carriers we carried out a number of crosses and quantified progeny genotype (Table 1). Stocks consisting of translocation homozygotes appeared generally healthy as adults, and survival from egg to adult was 96% of that observed for the Canton S (CS) wildtype stock. In contrast, crosses between males or females heterozygous for the translocation and wildtype resulted in semisterility, with only about 50% of progeny surviving to adulthood, and 50% of the survivors being translocation heterozygotes. These are the expected results if alternate and adjacent-1 segregation occur with equal frequency in translocation-bearing individuals during meiosis, resulting in the production of 50% aneuploid gametes (Figure 1B). Finally, for each translocation type we also carried out crosses between male and female translocation heterozygotes. Only 37.5% of progeny are predicted to survive, due to the large fraction of zygotes carrying unbalanced chromosome complements. However, many of the survivors (83%) are predicted to carry one or two copies of the translocation (Figure 1B). The levels of embryo survival and percentage of adults carrying the translocation were in good agreement with these predictions (Table 1). Together, these observations suggest that the translocation-bearing strains are fit (notwithstanding the expected semisterility), at least to a first approximation. These points notwithstanding, fitness measurements such as these are not sufficient to know that frequency-dependent drive will occur. This is well illustrated by the results of Curtis and Robinson, who found that a 2;3 translocation strain generated with X-rays, which had homozygous viability and fertility equivalent to wildtype in crosses such as those described above, was unable to drive population replacement, even when introduced at a 9:1 translocation:wildtype ratio (49).

For population replacement experiments we first introgressed our translocation-bearing systems, 51C; 70A2 and 51C; 68E flies, with Canton S (CS) for 8 generations, so as to minimize background genetic differences between translocation-bearing and wildtype strains. Translocation-bearing individuals

were then backcrossed to each other to create homozygous stocks. We initiated population cage experiments by introducing translocation-bearing males and virgin females into cages along with Canton S males and virgin females of similar age. A number of different introduction frequencies were tested, in triplicate. These included frequencies predicted to be super-threshold (80%, 70%, 60%), and sub-threshold (20%, 30%, 40%). Populations were then followed for 14 generations, with the frequency of translocation-bearing individuals noted each generation.

Results of these experiments are summarized in Figure 5A,B (solid lines). For both translocation-bearing strains, all nine releases at frequencies lower than 50% resulted in elimination of the translocation from the population. Conversely, introductions at frequencies greater than 50% resulted in translocation-bearing genotypes spreading to high frequency. These results are generally consistent with the modeling predictions. However, the dynamics of drive are clearly distinct from those predicted for translocations that lack a fitness cost (dotted lines in Figure 5A,B). When translocations were introduced at predicted super-threshold frequencies spread was slower than expected for a translocation with no fitness cost. Sub-threshold releases also resulted in lower initial translocation frequencies than expected, and this was generally followed in later generations by a modestly decreased time to elimination as compared with a translocation with no fitness cost (except at the 20% introduction frequency).

To understand these dynamics, we fitted the experimental data with our previously described deterministic model framework (19) using a range of different fitness cost models (Methods). By comparing the Akaike Information Criterion (AIC) values for each of these fitness cost models we found the best fitting model for the observed population dynamics to be one in which the relative fitness of homozygotes having the translocation is time-dependent, with the relative fitness of these individuals rapidly increasing over time, at first rapidly and converging upon some higher value as described by an exponential function.

Calculations of fitness parameters for translocation system 1 suggest an initial relative fitness of transgenic homozygotes of 0.0004 (95% CrI: 0-0.0019) relative to wild-types in generation 1 (the first progeny generation post adult introduction), rising to a relative fitness of 1.51 (95% CrI: 1.48-1.53) in subsequent generations. Calculations suggest an initial relative fitness of transgenic heterozygotes of 1.23 (95% CrI: 1.14-1.31) relative to wild-types, falling slightly to a relative fitness of 1.05 (95% CrI: 1.02-1.08). Calculations for translocation system 2 suggest an initial relative fitness of transgenic homozygotes of 0.0003 (95% CrI: 0-0.0016) relative to wild-types, rising to a relative fitness of 1.52 (95% CrI: 1.50-1.55) in subsequent generations, and an initial relative fitness transgenic heterozygotes that remains fairly constant: 1.12 (95% CrI: 1.05-1.18) at the beginning of the experiment and 1.11 (95% CrI: 1.08-1.14) at the end of the experiment.

While speculative, the initial very low fitness of homozygotes in generation 1 could reflect the fact that these individuals must derive from homozygous translocation parents. Our analysis of fitness presented in table 1 only examines viability, not ability to compete against other genotypes. Decreased fitness of homozygotes in competition with heterozygotes and wildtypes at some life stage (such as larval competition) could reflect incomplete removal of deleterious mutations during introgression into the CS background prior to carrying out drive experiments since recombination on translocation-bearing chromosomes in Drosophila is reduced throughout the involved arms (67, 68). Alternatively, it could also reflect the acquisition of genetic modifiers during the post-introgression crosses of the translocation stocks required to generate large numbers of homozygotes for population cage experiments. Such modifiers would, in this model, increase the fitness of homozygous carriers in competition with non-carrier homozygotes, but would result in a cost to carriers when in competition with heterozygotes and wildtypes. In either of these models it is unclear why fitness of translocations becomes greater than that of wildtype in later generations. Understanding the basis for these dynamics, and whether they

475 are specific to these translocations, will require further study in other genetic
476 backgrounds, and with other engineered translocations.
477

Discussion

Here we report the creation of engineered reciprocal translocations able to drive high threshold population replacement in Drosophila. The tools we used to create translocations in Drosophila - transgene cassettes located on two different chromosomes, a dominant marker created through the act of translocation, a site-specific nuclease able to bring about breakage within each cassette, and unique sequences that can mediate recombination between the two chromosomes - should be portable to other species. This, coupled with the common genetic behavior of reciprocal translocations in diverse species (semisterility in heterozygotes), suggests that translocation-based, high threshold and reversible drive may be possible in many species.

An important unknown from previous work is whether engineered translocations with high fitness are rare or common. Our observations demonstrating population replacement at high but not low introduction frequencies, while limited to two translocations sharing one breakpoint in common, suggest that engineered translocations with high fitness may at least not be rare. That said, while the translocations we generated are competitive in laboratory populations, it remains to be shown that these or any other engineered translocations are fit in competition with the diversity of genotypes that will be encountered in complex natural environments.

Our modeling results suggest that given high enough introduction frequencies, even translocations with high fitness costs, and facing significant levels of incoming migration of wildtypes, can spread to high frequency within a target area. However, modeling also identifies a set of tradeoffs associated with high threshold gene drive. Population replacement is local, but gene flow due to migration has significant effects on the equilibrium frequencies of transgenes within and outside the target area. Consideration of these effects will be important in identifying contexts in which population replacement is likely to have

509 an epidemiological impact, and is able to satisfy regulatory requirements relating
510 to the presence and movement of transgene-bearing organisms. These points on
511 gene flow within a target species notwithstanding, translocation-based drive
512 should be very species specific. This is because drive involves the behavior of
513 entire recombinant chromosomes. It seems unlikely that such a novel entity
514 would thrive when transferred to a different species through mating or horizontal
515 gene transfer.
516
517 A key feature of any population replacement mechanism is its degree of
518 evolutionary stability. A translocation drives because its presence in a single
519 copy in heterozygotes creates a toxic condition (genomic imbalance in some
520 gametes) that can be prevented by a second copy of the translocation, which
521 results in the creation of a fit translocation homozygote (genomic balance in all
522 gametes). One can think of this as a toxin-antidote system in which the toxin (the
523 translocation) is dominant (one copy results in genomic imbalance and some
524 death) and the antidote is recessive (two copies of the translocation results in
525 genomic balance and progeny viability). However, in contrast to other toxin-
526 antidote gene drive systems (23, 32, 33, 35-40), the toxin and antidote functions
527 of a translocation are inextricably linked: the toxin is the translocation (in one
528 copy), and the antidote is also the translocation (in two copies). It is presumably
529 very unlikely that the translocation will revert back to the wildtype chromosome
530 configuration. However, even if this happened, necessarily in a single rare
531 individual, this chromosome would be eliminated along with other wildtype
532 chromosomes in a population (of this or any other species (see above)) in which
533 the translocation was present at high frequency. In short, translocation-
534 dependent gene drive cannot break down through mutation of toxin function to
535 inactivity, as with many other chromosomally based drive mechanisms. It is also
536 insensitive to chromosomal sequence variation, mutation and non-homologous
537 end joining, which can prevent the spread of homing-based gene drive
538 mechanisms that rely on cleavage of a specific target sequence (69, 70). Finally,
539 the genes of interest will be placed at the translocation breakpoints. Meiotic

recombination is inhibited in these regions (67, 68). In addition, the transgenes are not located in regions that undergo pairing during meiosis. Since they are insertions of novel sequences, they are adjacent to regions that undergo pairing. Thus, transgenes are unlikely to become unlinked from the translocation breakpoint.

Finally, with any population replacement strategy one must plan for the eventual failure of the cargo, whether it encodes one or more genes that mediate disease resistance, or conditional lethality. Failure can occur through evolution of the pathogen. It can also occur through mutational inactivation of the cargo genes. In this latter case, if loss of cargo gene function also results in loss of an associated fitness cost, chromosomes carrying the mutant allele will spread at the expense of those carrying the functional allele. While mutation to inactivity cannot be prevented, chromosome-based drive mechanisms such as translocations have the attractive feature that it should be possible to incorporate multiple transgenes near the breakpoints, bringing about redundancy in effector function and thereby increased functional lifetime in the wild. Cycles of population replacement to bring new genes into the population can also be imagined. In one approach, the translocation can first be removed from the population by driving its frequency below the threshold needed for drive, through dilution with wildtypes. This can then be followed by a second release of a new translocation-bearing strain that has the same breakpoints, and a new cargo. Alternatively, if high fitness translocations with distinct breakpoints can be generated routinely, it may be possible to drive a first generation translocation and any remaining widltypes out of the population in favor of a second, distinct translocation (a point also made by Serebrovskii (47) in the context of use of translocations for population suppression) carrying a new cargo, as with proposals for cycles of replacement of *Medea*-based gene drive systems (5, 21).

The above positive points notwithstanding, several unknowns remain to the implementation of translocation-based population replacement in other insects.

First, generating translocations with the approaches described herein will be more challenging in other species in which a high quality annotated genome sequence is not available. Such a resource allows one to identify gene deserts, good candidates for sites in which to locate breakpoints associated with a minimal fitness cost to carriers. It also allows one to determine the orientation with respect to the centromere of sequences that mediate homologous recombination at breakpoints, so as to promote the formation of translocations rather than dicentric and acentric chromosomes. As an example, while the level of annotation of the *Aedes aegypti* genome sequence and transcriptome is otherwise quite high, much of the genome is annotated as a series of contigs of unknown orientation, due to the large amount of repetitive sequences in the genome. Finally, a sequenced genome makes it possible to identify or create, using HEGs, Zinc fingers, TALENs or Crispr/Cas9, site-specific nucleases that promote recombination by cleaving within the transgenes but not elsewhere in the genome.

In addition, the models we have used to characterize translocation behavior do not take into account important real world variables such as non-random mating and local spatial heterogeneity, which can affect the dynamics of translocation spread (55, 71). In order to understand how these and other environmental variables effect translocation-based replacement, and high threshold replacement more generally, it will be important to model drive element behavior using spatially explicit models based on analysis of real populations in complex environments (72, 73). Finally, mosquito populations in the wild consist of multiple chromosomal forms, and may also display some level of reproductive isolation (74-76). How engineered translocations will fare in the face of these variants remains to be determined, but can be explored in competition with genetically diverse laboratory strains (77, 78). While an understanding of the above issues is critical for the success of any population-replacement strategy, the problems are not intractable, as evidenced by successes in controlling pest

601 populations using non-transgenic (79) and transgenic inundative population
602 suppression strategies (80, 81).

603 **Methods**

604 **Construct Assembly**

605 The Gibson enzymatic assembly (EA) cloning method was used for all cloning
606 (82). For both constructs (A and B), translocation allele components were cloned
607 into the multiple cloning site (MCS) of a plasmid (83) containing the *white* gene
608 as a marker and an attB-docking site. For construct A (Figure 1B), the oenocyte-
609 specific *svp* enhancer (65) and Hsp70 basal promoter fragments were amplified
610 from *Drosophila melanogaster* genomic DNA using primers P16 and P17 (*svp*)
611 and P18 and P19 (Hsp70). The GFP fragment was amplified from template
612 pAAV-GFP (addgene plasmid #32395) using primers P26 and P27. A Kozak
613 sequence (CAACAAA) directly 5' of the GFP start codon was added with primer
614 P26. The SV40 3'UTR fragment was amplified from template pMos-3xP3-DsRed-
615 attp (addgene plasmid #52904) using primers P28 and P10. The 5' and 3' CTCF
616 insulator fragments (84) were amplified from *Drosophila melanogaster* genomic
617 DNA using primers P11 and P15 (for the 5' CTCF fragment) and P13 and P14
618 (for the 3' CTCF fragment). The 667 XYZ and 668 UVW homology fragments
619 were amplified as above with primers P22 and P23 (XYZ) and P20 and P21
620 (UVW), from plasmid pFUSE-mIgG1-Fc Invivogen, San Diego). The 5' and 3'
621 splice sites utilized were from a 67bp intron located in the *Drosophila*
622 *melanogaster* Myosin Heavy Chain (Mhc) gene ID CG17927. They were added
623 to UVW and XYZ sequences using PCR; the 5' splice site was added to the 5'
624 end of the UVW fragment via PCR with primer P24, and the 3' splice site was
625 added to the 3' end of fragment XYZ via PCR with primer P25. Two I-SceI
626 recognition sequences Two 18bp I-SceI recognition sequences
627 (ATTACCCTGTTATCCCTA-CTAG-TAGGGATAACAGGGTAAT) were added to
628 the 3' end of the UVW fragment with primer P21 and the 5' end of the XYZ
629 fragment with primer P22. The construct was assembled in two steps, as above,
630 with the first (5') CTCF, the svp and hsp70 fragments, the UVW fragment, and
631 the XYZ fragment cloned in via a first EA cloning step, and the GFP fragment,
632 the SV40 3'UTR fragment, and the second (3') CTCF cloned in via a second EA
633 cloning step. For construct B (Figure 1B), the opie2 promoter fragment was

amplified from plasmid pIZ/V5-His/CAT (Invitrogen) using primers P1 and P2. The XYZ and UVW homology fragments were amplified from plasmid pFUSEss-CHIg-mG1 using primers P3 and P4 (XYZ) and P5 and P6 (UVW). Two 18bp I-SceI recognition sequences (ATTACCCTGTTATCCCTA-CTAG-TAGGGATAACAGGGTAAT) were added to the 3' end of the XYZ fragment and the 5' end of the UVW fragment in inverse orientation to each other separated by a 4bp linker sequence (CTAG) using primers P4 (for XYZ) and P5 (for UVW). The 5' and 3' splice sites utilized were from a 67bp intron located in the *Drosophila melanogaster* Myosin Heavy Chain (Mhc) gene ID CG17927; the 5' splice site was added to the 5' end of the XYZ fragment via PCR with primer P7, and the 3' splice site was added to the 3' end of fragment UVW via PCR with primer P8. The dsRed fragment, together with the SV40 3'UTR, were amplified from template pMos-3xP3-DsRed-attp (addgene plasmid #52904) using primers P9 and P10, with a Kozak sequence (CAACAAA) directly 5' of the DsRed start codon added with primer P9. The 5' and 3' CTCF insulator fragments (84) were amplified from *Drosophila melanogaster* genomic DNA using primers P11 and P12 (for the 5' CTCF fragment) and P13 and P14 (for the 3' CTCF fragment). The construct was assembled in two steps. First, the *Drosophila melanogaster* attB stock plasmid (83) was digested with AscI and XbaI, and the first (5') CTCF, the opie-2 promoter, the XYZ fragment, and the UVW fragments were cloned via EA cloning. Then, the resulting plasmid was digested with XhoI, and the dsRed-SV40 3'UTR fragment and the second (3') CTCF were cloned in via EA cloning. All sequences were analyzed with NNSPLICE 0.9 (available at http://www.fruitfly.org/seq_tools/splice.html) to confirm strength of splice signals and to check for cryptic splice sites. A list of primer sequences used in the above construct assembly can be found in Supplementary Table 1.

Fly Culture and Strains

Fly husbandry and crosses were performed under standard conditions at 25°C. Rainbow Transgenics (Camarillo, CA) carried out all of the fly injections. Bloomington Stock Center (BSC) fly strains utilized to generate translocations

were attP lines 68E (BSC #24485: y^1 M{vas-int.Dm}ZH-2A w^*; M{3xP3-RFP.attP'}ZH-68E), 51C (BSC #24482; y[1] M{vas-int.Dm}ZH-2A w[*]; M{3xP3-RFP.attP'}ZH-51C), and 70A2 (BSC #9741: y[1] w[1118]; PBac{y[+]-attP-9A}VK00023). Fly Stock BSC#6935 (y[1] w[*]; P{ry[+t7.2]=70FLP}23 P{v[+t1.8]=70I-SceI}4A/TM) was used as the source of heat shock induced I-SceI. For balancing chromosomes, fly stocks BSC#39631 (w[*]; wg[Sp-1]/CyO; P{ry[+t7.2]=neoFRT}82B lsn[SS6]/TM6C, Sb[1]) BSC#2555 (CyO/sna[Sco]) were used. For introgression into a wild type background we used the Canton-S stock BSC#1. Translocation construct A was inserted at site 51C, and construct B was inserted at 68E and 70A2 using phiC31 mediated attP/attB integration. These site combinations allowed for the generation of two distinct translocation types, 51C;68E and 51C;70A2. Stocks homozygous for both constructs were then mated with flies that express I-SceI under the control of the Hsp70 heat shock promoter(66). Progeny carrying all three transgenes were subjected to 5 rounds of heat shock during larval stages and as adults. Heat shocks were conducted by submerging fly vials in a water bath set to 38°C for one hour. Adults were outcrossed to w-, and progeny examined under a fluorescent dissecting scope for ubiquitous GFP expression, indicative of translocation generation.

Homozygous translocation-bearing stocks were generated for both 51C;68E and 51C;70A2 site combinations by crossing translocation heterozygotes and identifying homozygous progeny by eye color (light orange eyes for homozygotes versus yellow for heterozygotes for the 51C;68E site combination; light red eyes for homozygotes versus orange for heterozygotes for the 51C;70A2 site combination. After confirming homozygous viability, translocations were introgressed into a Canton-S genetic background. First, CS females were crossed to translocation-bearing males so as to bring the CS mitochondrial genotype into the translocation background. Subsequently, translocation heterozygote females were outcrossed to CS males for 8 generations. Heterozygous translocation-bearing males and virgin females were then crossed to each other to generate homozygous stocks in the CS background for each site

combination. Homozygosity was confirmed by outcrossing. Drive experiments for these stocks were set up against CS as the wildtype stock.

Embryo and Adult viability determination

For embryo viability counts (Table 1), 2-4 day old adult virgin females were mated with males of the relevant genotypes for 2-3 days in egg collection chambers, supplemented with yeast paste. On the following day, a 3hr egg collection was carried out, after first having cleared old eggs from the females through a pre-collection period on a separate plate for 3hrs. Embryos were isolated into groups and kept on an agar surface at 25^{o}C for 48-72 hrs. The % survival was then determined by counting the number of unhatched embryos. One group of 100-200 embryos per cross was scored in each experiment, and each experiment was carried out in biological triplicate. The results presented are averages from these three experiments. Embryo survival was normalized with respect to the % survival observed in parallel experiments carried out with the Canton-S wild-type strain, which was 93.00% $\pm$ 1.82%. For adult fly counts (Table 1), individual flies for each genotype cross were singly mated. For each genotype cross, we set up 10-15 individual fly crosses, and the results presented are averages from all these experiments.

Population cage experiments

All population cage experiments were carried out at 25^{o}C, 12 hour-12 hour day night cycle, with ambient humidity in 250 ml bottles containing Lewis medium supplemented with live, dry yeast. Starting populations for drive experiments included equal numbers of virgins and males of similar ages, for each genotype. Translocation-bearing homozygotes were introduced at population frequencies of 60%, 70%, and 80% (T_1/T_1; T_2/T_2) for above threshold drive experiments, and 20%, 30%, and 40% (T_1/T_1; T_2/T_2) for below threshold drive experiments. CS virgin females and males (+/+; +/+) of similar age as the translocation-bearing individuals made up the remainder of the population. The total number of flies for each starting population was 100. All experiments were

conducted in triplicate. After being placed together, adult flies were removed after seven days. After another seven days, progeny were collected and divided arbitrarily into two equally sized groups. For one group the fraction of translocation-bearing individuals (T_1/T_1; T_2/T_2 or $T_1/+$; $T_2/+$) was determined, while the other group was placed into a new bottle to initiate the next generation.

Theoretical Framework

We apply the model of Curtis and Robinson (1971) to describe the spread of reciprocal translocations through a population. This is a discrete-generation, deterministic population frequency model assuming random mating and an infinite population size. We denote the first chromosome with a translocated segment by "*T*" and the wild-type version of this chromosome by "*t*." Similarly, we denote the second chromosome with a translocated segment by "*R*" and the wild-type version of this chromosome by "*r*." As a two-locus system, there are nine possible genotypes; however, only individuals carrying the full chromosome complement are viable, which corresponds to the genotypes *TTRR*, *TtRr* and *ttrr*, the proportion of the *k*th generation of which are denoted by p_k^{TTRR}, p_k^{TtRr} and p_k^{ttrr}. The four haplotypes that determine the genotype frequencies in the next generation – *TR*, *tR*, *Tr* and *tr* – are described by the following frequencies:

$$f_k^{TR} = p_k^{TTRR}(1-s) + 0.25\,p_k^{TtRr}(1-hs)$$

$$f_k^{tR} = f_k^{Tr} = 0.25\,p_k^{TtRr}(1-hs)$$

$$f_k^{tr} = p_k^{ttrr} + 0.25\,p_k^{TtRr}(1-hs)$$

Here, *s* denotes the reduced fecundity of *TTRR* individuals and *hs* denotes the reduced fecundity of *TtRr* individuals relative to wild-type individuals, where $h \in [0,1]$. By considering all possible mating pairs, the genotype frequencies in the next generation are:

$$p_{k+1}^{TTRR} = (f_k^{TR})^2 / \sigma_k$$

$$p_{k+1}^{TtRr} = 2(f_k^{TR} f_k^{tr} + f_k^{tR} f_k^{Tr}) / \sigma_k$$

$$p_{k+1}^{trrr} = (f_k^{tr})^2 / \sigma_k$$

where σ_k is a normalizing term given by,

$$\sigma_k = (f_k^{TR})^2 + 2(f_k^{TR} f_k^{tr} + f_k^{tR} f_k^{Tr}) + (f_k^{tr})^2$$

For our three-population models, there are three sets of the above equations to represent each population. We let *m* represent the migration rate per generation. After genotype frequencies for all three populations are calculated for a given generation, a proportion *m* is removed from each genotype from populations 1 and 3 and added to population 2, and a proportion 2*m* is removed from each genotype from population 2, half of which is added to population 1 and the other half of which is added to population 3.

We investigated a number of different fitness cost models and chose the one that provided the best fit to the data. In all cases, the parents in the first generation were not subject to a fitness cost. The simplest model is one in which the fitness of each genotype stays constant over time. Another model considers fitness costs that depend on the population frequency of the genotype. For linear frequency-dependence, this is given by,

$$s = (s_0 - s_1) p_k^{trrr} + s_1$$

Here, s_0 represents the fitness cost of a translocation homozygote in an almost fully wild-type population, and s_1 represents the fitness cost in an almost fully transgenic population. An alternative model is that fitness is time-dependent, as could be explained by introgression of introduced genotypes. For linear time-dependence, this is given by,

$$s = \left(\frac{s_1 - s_0}{t_f} \right) t + s_0$$

Here, s_0 represents the fitness cost in the second generation and s_1 represents the fitness cost in the final generation, denoted by t_f. For sigmoidal time-

780 dependence, it is given by,

781 $$s = (s_0 - s_1)\left(1 - \frac{1}{1 + e^{-\alpha(1-\tau)}}\right) + s_1$$

782 Here, s_0 and s_1 are as before, τ denotes the time of intermediate fitness cost,
783 and α denotes the speed of transition between the two fitness costs.

784 And for exponential time-dependence, it is given by,

785 $$s = a2^{-t/t_{1/2}} + (s_0 - a)$$

786 Here, s_0 represents the fitness cost in the second generation, s_1 represents the
787 fitness cost after many generations, $t_{1/2}$ denotes the time at which the fitness cost
788 is halfway between the two, and *a* is given by,

789 $$a = \frac{s_0 - s_1}{1 - 2^{-t_f/t_{1/2}}}$$

790 We estimated fitness parameters for each model and compared models
791 according to their Akaike Information Criterion (AIC) values. Model fitting was
792 performed using population count data for the 18 drive experiments conducted
793 for each translocation system (three for each of the 80%, 70%, 60%, 40%, 30%
794 and 20% release frequencies). AIC was calculated as $2k - 2\log L$, where k
795 denotes the number of model parameters, and the preferred model is the one
796 with the smallest AIC value. The likelihood of the data was calculated, given
797 fitness costs *s* and *hs*, assuming a binomial distribution of the two phenotypes
798 (individuals homozygous or heterozygous for the translocation were considered
799 as the same phenotype to match the experimental counts). Model predictions
800 were used to generate expected genotype proportions over time for each fitness
801 cost, and the log likelihood had the form,

802 $$\log L(h,s) = \sum_{i=1}^{18}\sum_{k=1}^{14} \log\binom{TTRR_{i,k} + TtRr_{i,k} + ttrr_{i,k}}{TTRR_{i,k} + TtRr_{i,k}} + ttrr_{i,k}\log(p_{i,k}^{ttrr}(h,s)) + (TTRR_{i,k} + TtRr_{i,k})\log(1 - p_{i,k}^{ttrr}(h,s))$$

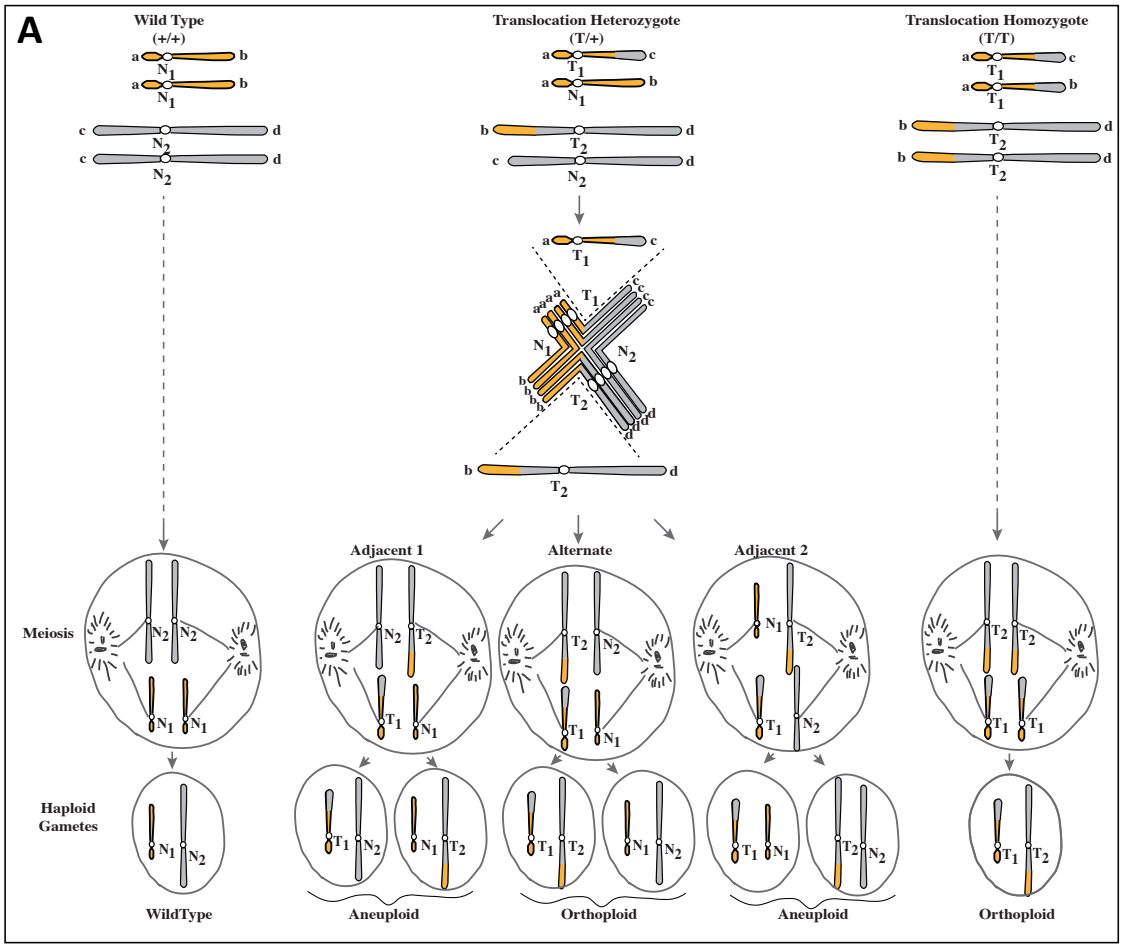
803 Here, $TTRR_{i,k}$, $TtRr_{i,k}$ and $ttrr_{i,k}$ represent the number of *TTRR*, *TtRr* and *ttrr*
804 individuals at generation *k* in experiment *i*, and the corresponding expected
805 genotype frequencies are fitness cost-dependent. The best estimate of the
806 fitness cost is that having the highest log-likelihood. A 95% credible interval was
807 estimated using a Markov Chain Monte Carlo sampling procedure. Matlab and R
808 code implementing these equations is available upon request. The AIC values for
809 each of the fitness cost models are shown in the table below:

Fitness cost model:	**AIC (Translocation system 1):**	**AIC (Translocation system 2):**
Constant fitness costs	6577.6	7448.0
Linear, frequency-dependent fitness costs	5051.3	5572.7
Linear, time-dependent fitness costs	3888.2	3752.1
Sigmoidal, time-dependent fitness costs	3344.2	3321.1
Exponential, time-dependent fitness costs	3336.2	3319.1

810

811 In summary, the best fitting model for the observed population dynamics is one in
812 which the relative fitness of homozygotes having the translocation is time-
813 dependent, with the relative fitness of these individuals increasing over time, at
814 first rapidly and then converging upon some higher value as described by an
815 exponential function (Figure 5).

816



817

818 **Figure 1**. Gamete and zygote genotypes associated with the presence of a
819 reciprocal translocation. Wildtype chromosomes N_1 and N_2, and translocation
820 chromosomes T_1 and T_2, are indicated. (A) One chromosome type (a) is
821 indicated in yellow. A second chromosome type (b) is in gray. Gamete types
822 generated by wildtype (+/+), translocation heterozygotes (T/+), and translocation
823 homozygotes (T/T) are indicated. (B) Gamete and zygote genotypes possible in

824 crosses involving a translocation are indicated. Inviable genotypes are indicated
825 by a red line.
826
827
828

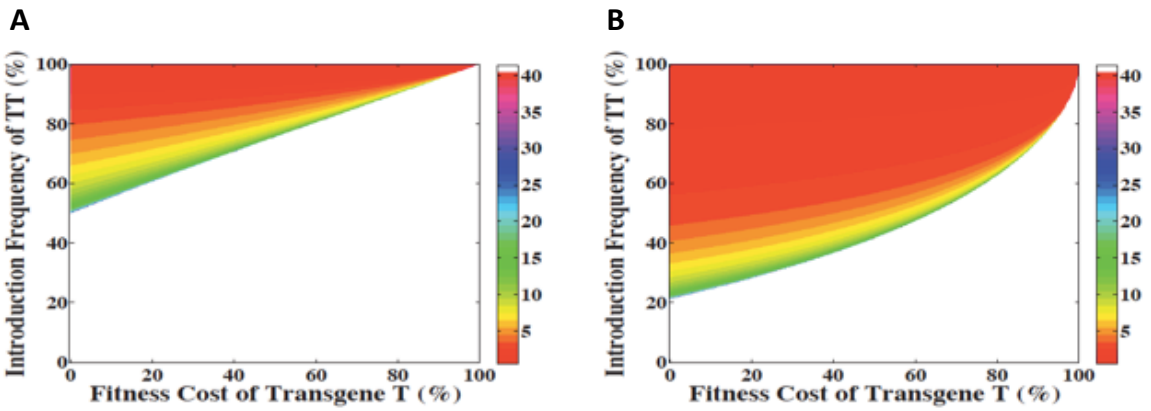


Figure 2. Engineered reciprocal translocations are predicted to show threshold-dependent gene drive and bring about local population replacement. A discrete generation, deterministic population frequency model of translocation spread through a single population for varying introduction frequencies and fitness costs for one (A) or three (B) introductions at the specified frequency. The heatmap indicates the number of generations required for the translocation to reach fixation (i.e., >99% of the total population) for all combinations of fitness cost and introduction frequency.

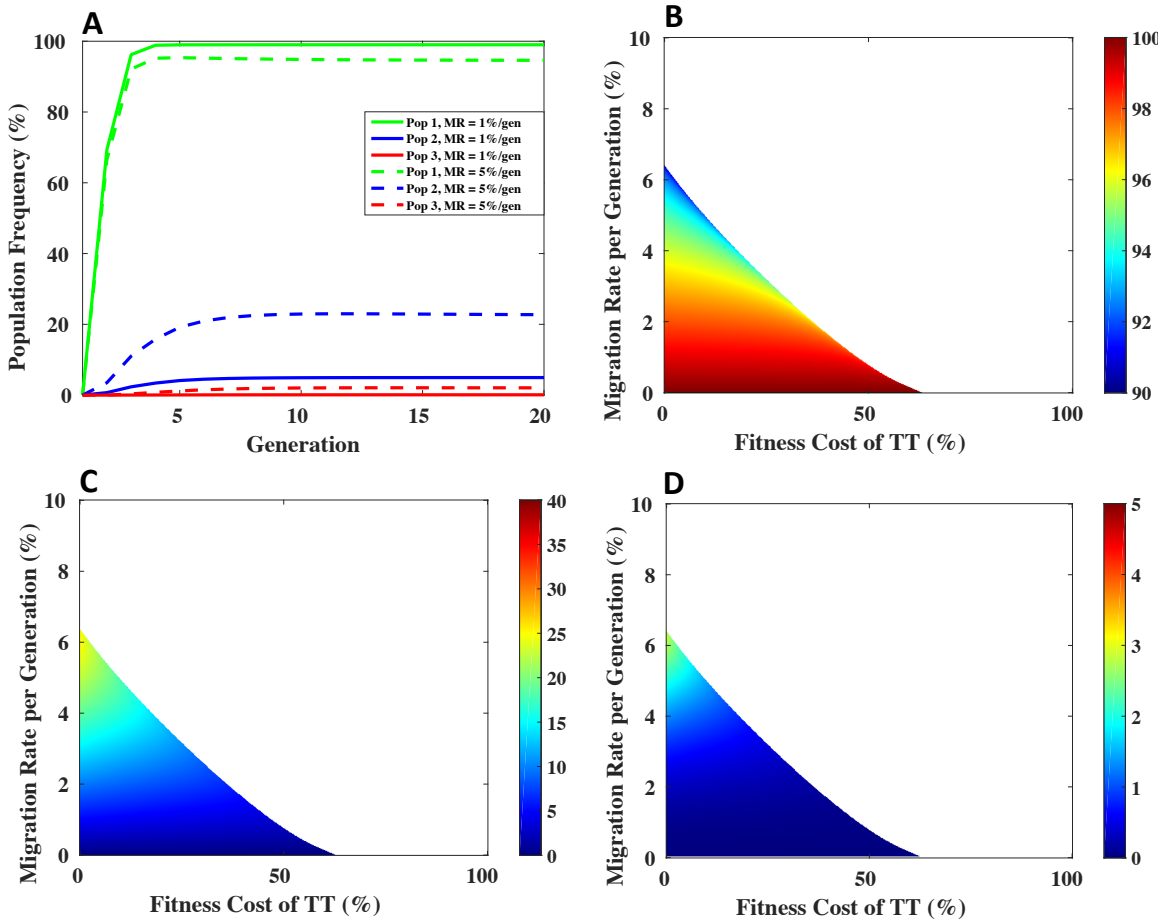


Figure 3. Translocation dynamics in a linear, three population migration model. (A) Population frequency of a translocation with no fitness cost, introduced into population 1 using three consecutive releases of translocation-bearing homozygotes. Populations 1-3 are linked through a linear chain of migration of 1% (solid lines) or 5% (dashed lines). (B-D) Equilibrium frequency of translocation bearing individuals over a range of fitness costs and migration rates for each of the three linked populations 1 (B), 2 (C), and 3 (D), respectively. For all three populations increasing fitness cost has little effect on the equilibrium frequency at low migration rate and increased effects at higher migration rates. In contrast, migration rate has a much stronger effect on equilibrium frequency independent of fitness cost as seen by the color gradient shifts. Note that the equilibrium frequency varies between 90-100%, 0-25%, and 0-3% in the target population (population 1), population 2, and population 3, respectively.

859 **Figure 4**

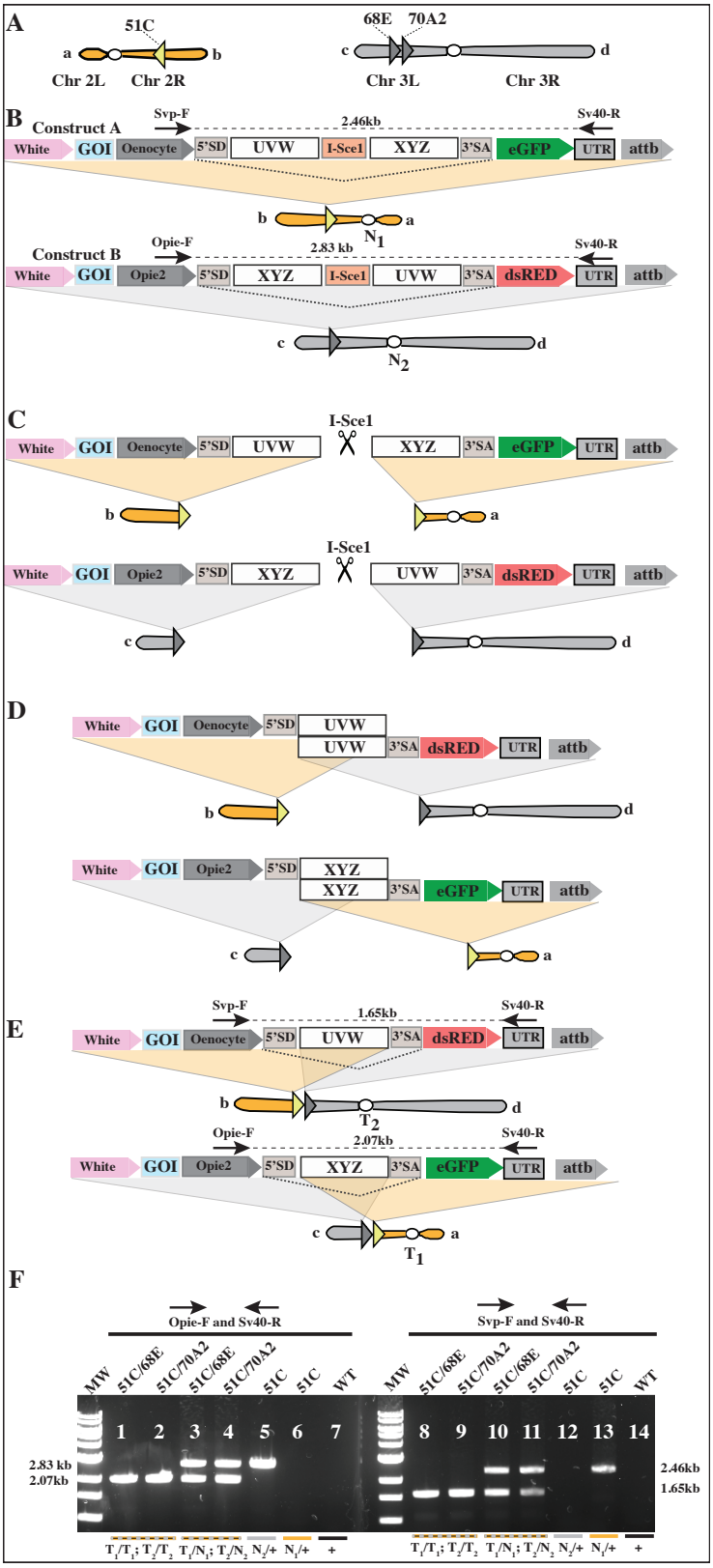


Figure 4. Generation of reciprocal translocations in Drosophila. (A) Approximate location of the attP sites used for transgene insertion; orientation with respect to

the centromere are indicated by triangles. (B) Components of each starting transgene cassette. Construct A is inserted on the second chromosome and construct B on the third chromosome. Components are as indicated in the text. (C) I-Sce-dependent cleavage results in a double-stranded break in each transgene-bearing chromosome. (D) Alignment of broken chromosome ends occurs using homologous sequences UVW and XYZ. (E) Recombinant chromosomes are generated by homologous recombination using sequences UVW and XYZ. (F) Agarose gel image is shown of PCR amplification products generated from different genotypes: translocation homozygotes (T1/T1; T2T2); translocation heterozygotes (T1N1; T2N); individuals carrying only the 51C starting chromosome insertion (N1/+); or the 68E and 70A2 starting chromosome insertion (N2/+). Primers used, and expected amplification product sizes, are indicated in B and E.

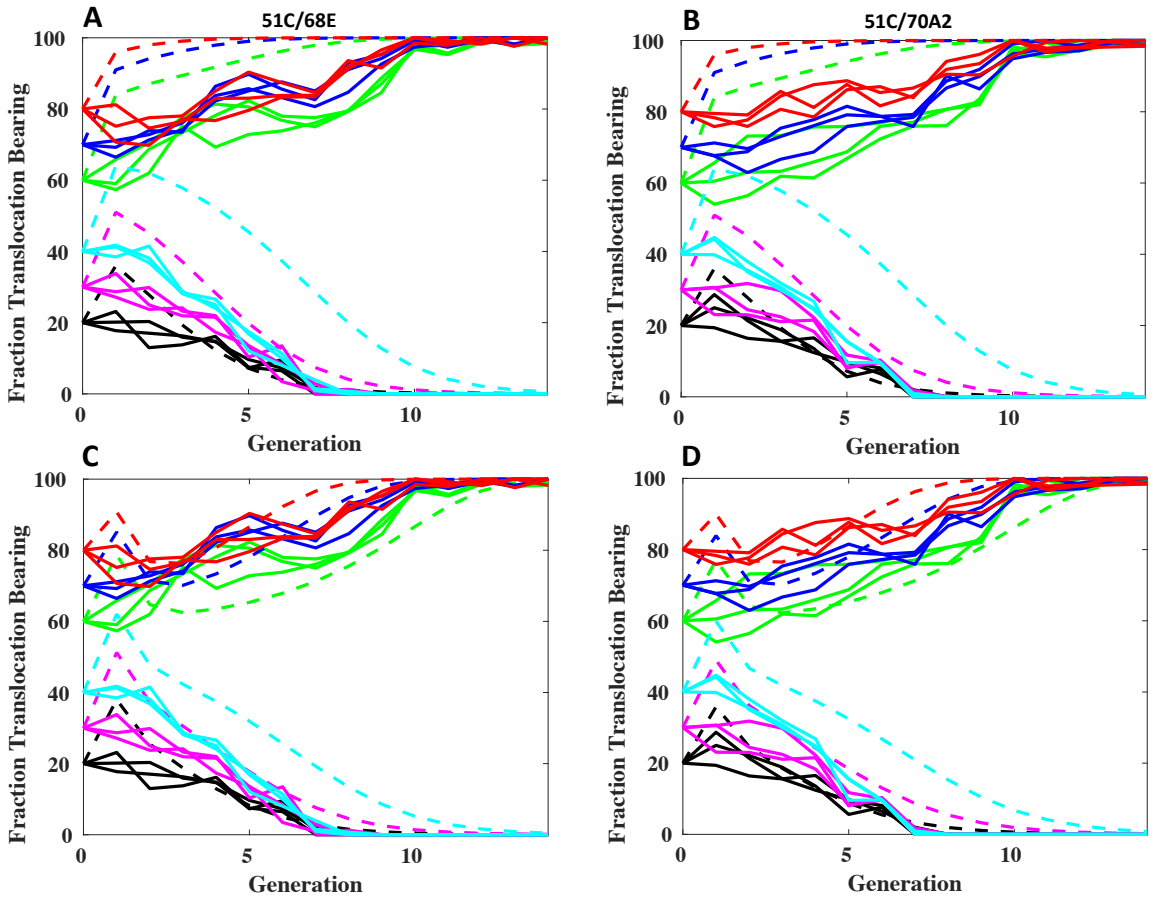


Figure 5. Dynamics of translocation-based population replacement, and predictions from zero fitness cost, and best fit models. (A, B) Population frequency of the adult population having the indicated translocation is plotted versus generation number for a number of homozygous translocation release ratios: 80%, 70%, 60%, 40%, 30% and 20%. Solid lines indicate observed population frequencies, and dashed lines indicate predicted translocation-bearing genotype frequencies for an element with no fitness cost. (C, D). The same data as in (A, B) but plotted along with dynamics predicted based on a best fit model described in the methods and text.

parental genotypes			embryo survival %		transgene bearing adults %	
male	female	progeny genotype (%)	predicted	observed*	predicted	observed*
T_1/T_1 ; T_2/T_2	T_1/T_1 ; T_2/T_2	T_1/T_1 ; T_2/T_2 (100%)	100	96.9 ± 1.8 96.9 ± 0.3	100	100 ± 0.0 100 ± 0.0
T_1/T_1 ; T_2/T_2	+/+ ; +/+	$T_1/+$; $T_2/+$ (100%)	100	94.6 ± 2.2 98.2 ± 2.6	100	100 ± 0.0 100 ± 0.0
+/+	T_1/T_1 ; T_2/T_2	$T_1/+$; $T_2/+$ (100%)	100	90.1 ± 1.6 92.5 ± 4.8	100	100 ± 0.0 100 ± 0.0
$T_1/+$; $T_2/+$	+/+ ; +/+	$T_1/+$; $T_2/+$ (25%) $T_1/+$; +/+ (25%)** +/+ ; $T_2/+$ (25%)** +/+ ; +/+ (25%)	50	51.2 ± 1.6 50.4 ± 1.3	50	49.3 ± 3.4 49.5 ± 2.4
+/+ ; +/+	$T_1/+$; $T_2/+$	$T_1/+$; $T_2/+$ (25%) $T_1/+$; +/+ (25%)** +/+ ; $T_2/+$ (25%)** +/+ ; +/+ (25%)	50	48.3 ± 2.8 48.3 ± 3.9	50	49.4 ± 2.2 48.5 ± 3.4
$T_1/+$; $T_2/+$	$T_1/+$; $T_2/+$	T_1/T_1 ; T_2/T_2 (6.25%) T_1/T_1 ; $T_2/+$ (12.5%)** T_1/T_1 ; +/+ (6.25%)** $T_1/+$; T_2/T_2 (12.5%)** $T_1/+$; $T_2/+$ (25%) $T_1/+$; +/+ (12.5%)** +/+ ; T_2/T_2 (6.25%)** +/+ ; $T_2/+$ (12.5%)** +/+ ; +/+ (6.25%)	37.5	36.2 ± 1.8 32.4 ± 4.0	~83%	80.4 ± 6.5 80.8 ± 5.8

* Translocation 51C/68E (top) and 51C/9741 (bottom)
** These genotypes are not viable.

Table 1. Behavior of translocations in crosses to various genotypes. Crosses between parents of specific genotypes - wild-type (+/+; +/+), translocation heterozygotes ($T_1/+$; $T_2/+$), and translocation homozygotes (T_1/T_1; T_2/T_2), were carried out. Embryo survival (fifth column from right) and percentage of translocation-bearing adults (rightmost column) were independently quantified. The top number in each column shows results for the 51C/68E translocation; the bottom number shows the results for the 51C/70A2 translocation. ** Indicates unviable genotypes. Embryo survival was normalized with respect to percent survival (± SD) observed in the w^{1118} stock used for transgenesis (methods).

Supplementary Table 1. List of primer sequences used in this study.

Primer name	Primer sequence, 5' to 3'	Source
P1	CCTAACAACTCACACCTTGCAGCGCCACCTG	pIZ/V5-

P2	GCCCTAGAGATCCACCAACTTTTTTGCACTGC ATTCCTAAGCATCAGTGGTTGAACCTACCTTGTTGGCGTGACCAGAGACAGGTTGCGGCG	His/CAT (Invitrogen)
P3 P4	AGGTTCAACCACTGATGCTTAGGAATAGGCCATGTGAAGCTGAAGGAATC TATTACCCTGTTATCCCTACTAGTAGGGATAACAGGGTAATACTAGAATCCCTGGGCACAATTT	pFUSEss-CHIg-mG1 (Invivogen)
P5 P6	CTAGTATTACCCTGTTATCCCTACTAGTAGGGATAACAGGGTAATAGTGGTTGTAAGCCTTGCA AAAGGATAAGAATTAGGGTTAGTCGTTTCGGTGTGCCTAGTTTACCAGGAGAGTGGGAGA	pFUSEss-CHIg-mG1 (Invivogen)
P7	CGCCCACGCCATCCAACCGCCGCCGCAACCTGTCTCTGGTCACGCCAACAAGGTAGGTTC	P3/P4 XYZ PCR
P8	ATGACGTTCTTGGAGGAGCGCACCATTTTGTTGCTAAAGGAAAGGATAAGAATTAGGGTT	P5/P6 UVW PCR
P9 P10	AAACGACTAACCCTAATTCTTATCCTTTCCTTTAGCAACAAAATGGTGCGCTCCTCCAAG AATGGAACTCTTCGCGGCCAGGTGGCGCTGCAAGGCTCGAGGGTCGACTGATCATAATCA	pMos-3xP3-DsRed-attp (addgene plasmid #52904)
P11 P12	GGATCCGGGAATTGGGAATTGGGCAATATTTAAATGGCGGCCTTGCAGCGCCACCTGGCC AGCGTGTTTTTTTGCAGTGCAAAAAAGTTGGT	Drosophila genomic DNA

	GGATCTCTAGGGCCAGGTGGCGCTGCAA	
P15	CCAACGCATTTTCCAAGCTTGTTTAAACGTGG ATCTCTAGGGCCAGGTGGCGCTGCAAGG	
P13	TACAAATGTGGTATGGCTGATTATGATCAGTC GACCCTCGAGCCTTGCAGCGCCACCTGG	Drosophila genomic DNA
P14	GAGACCGTGACCTACATCGTCGACACTAGTG GATCTCTAGGGCCAGGTGGCGCTGCAAGG	
P16	CCTTGCAGCGCCACCTGGCCCTAGAGATCCA CGTTTAAACAAGCTTGGAAAATGCGTTGG	Drosophila genomic DNA
P17	CGAAGCGCCTCTATTTATACTCCGGCGCTCG TTTAAACAAAGTGGCAGGGCCCATGTGTT	
P18	GAGTGGAGCACAAACACATGGGCCCTGCCA CTTTGTTTAAACGAGCGCCGGAGTATAAAT	Drosophila genomic DNA
P19	AAGCATCAGTGGTTGAACCTACCTTGTTGGC GTGTCTGATGCAGATTGTTTAGCTTGTTC	
P20	GCCAACAAGGTAGGTTCAACCACTGATGCTT AGGAATAGGCGTGGTTGTAAGCCTTGCAT	pFUSEss-CHIg-mG1 (Invivogen)
P21	CCCTGTTATCCCTACTAGTAGGGATAACAGG GTAATACTAGTTTACCAGGAGAGTGGGAG	
P22	TATTACCCTGTTATCCCTACTAGTAGGGATAA CAGGGTAATACATGTGAAGCTGAAGGAA	pFUSEss-CHIg-mG1 (Invivogen)
P23	AAAGGATAAGAATTAGGGTTAGTCGTTTCGG TGTGCCTAGAATCCCTGGGCACAATTTTC	
P24	CAAGCGCAGCTGAACAAGCTAAACAATCTGC ATCAGACACGCCAACAAGGTAGGTTCAAC	P20/P21 UVW PCR

P25	ACCTACATCGTCGACACTAGTGGATCTCTAG CTCGAGCTAAAGGAAAGGATAAGAATTAGGG	P22/P23 XYZ PCR
P26 P27	CCCTAATTCTTATCCTTTCCTTTAGGAATTCC AACAAAATGGTGAGCAAGGGCGAGGAGC TTCACTGCATTCTAGTTGTGGTTTGTCCAAAC TCATCAATGTTTACTTGTACAGCTCGTC	pAAV-GFP (addgene plasmid #32395)
P28	GCCGCCGGGATCACTCTCGGCATGGACGAG CTGTACAAGTAAACATTGATGAGTTTGGAC	pMos-3xP3-DsRed-attp (addgene plasmid #52904)

Acknowledgements

Work in BAH's lab (BAH, OSA, ABB, and TI) was supported by the U. S. Army Research Laboratory and the U. S. Army Research Office under contract W911NF-11-2-0055 to the California Institute of Technology, and the USDA and CRDF. Work at UCR (OSA and AB) was supported by an NIH-K22 Career Transition award (5K22AI113060-02) and UCR lab startup funds to OSA. JMM was supported by funding from The Parker Foundation through a gift to the University of California, San Francisco, Global Health Group Malaria Elimination Initiative. TI was supported by NIH training grant 1432.

1. Nicholson GM (2007) Fighting the global pest problem: preface to the special Toxicon issue on insecticidal toxins and their potential for insect pest control. *Toxicon* 49(4):413-422.
2. Curtis CF (1968) Possible use of translocations to fix desirable genes in insect pest populations. *Nature* 218(5139):368-369.
3. Braig HR & Yan G (2001) The spread of genetic constructs in natural insect populations. *Genetically Engineered Organisms: Assessing Environmental and Human Health Effects*, eds Letourneau DK & Burrows BE (CRC Press), pp 251-314.
4. Sinkins SP & Gould F (2006) Gene drive systems for insect disease vectors. *Nat Rev Genet* 7(6):427-435.
5. Hay BA*, et al.* (2010) Engineering the genomes of wild insect populations: challenges, and opportunities provided by synthetic Medea selfish genetic elements. *J Insect Physiol* 56(10):1402-1413.
6. Burt A (2014) Heritable strategies for controlling insect vectors of disease. *Philos Trans R Soc Lond B Biol Sci* 369(1645):20130432.
7. Whitten MJ (1971) Insect control by genetic manipulation of natural populations. *Science* 171(3972):682-684.
8. Akbari OS*, et al.* (2014) Novel synthetic Medea selfish genetic elements drive population replacement in Drosophila; a theoretical exploration of Medea-dependent population suppression. *ACS Synth Biol* 3(12):915-928.
9. Okamoto KW, Robert MA, Lloyd AL, & Gould F (2013) A reduce and replace strategy for suppressing vector-borne diseases: insights from a stochastic, spatial model. *PLoS One* 8(12):e81860.
10. Robert MA, Okamoto K, Lloyd AL, & Gould F (2013) A reduce and replace strategy for suppressing vector-borne diseases: insights from a deterministic model. *PLoS One* 8(9):e73233.
11. Okamoto KW, Robert MA, Gould F, & Lloyd AL (2014) Feasible introgression of an anti-pathogen transgene into an urban mosquito population without using gene-drive. *PLoS Negl Trop Dis* 8(7):e2827.
12. Robert MA, Okamoto KW, Gould F, & Lloyd AL (2014) Antipathogen genes and the replacement of disease-vectoring mosquito populations: a model-based evaluation. *Evolutionary applications* 7(10):1238-1251.
13. Knols BG, Bossin HC, Mukabana WR, & Robinson AS (2007) Transgenic mosquitoes and the fight against malaria: managing technology push in a turbulent GMO world. *Am J Trop Med Hyg* 77(6 Suppl):232-242.
14. Marshall JM (2010) The Cartagena Protocol and genetically modified mosquitoes. *Nat Biotechnol* 28(9):896-897.
15. Mumford JD (2012) Science, regulation, and precedent for genetically modified insects. *PLoS Negl Trop Dis* 6(1):e1504.
16. Alphey L & Beech C (2012) Appropriate regulation of GM insects. *PLoS Negl Trop Dis* 6(1):e1496.
17. Reeves RG, Denton JA, Santucci F, Bryk J, & Reed FA (2012) Scientific standards and the regulation of genetically modified insects. *PLoS Negl Trop Dis* 6(1):e1502.

18. Marshall JM (2009) The effect of gene drive on containment of transgenic mosquitoes. *J Theor Biol* 258(2):250-265.
19. Marshall JM & Hay BA (2012) Confinement of gene drive systems to local populations: a comparative analysis. *Journal of Theoretical Biology* 294:153-171.
20. Wade MJ & Beeman RW (1994) The population dynamics of maternal-effect selfish genes. *Genetics* 138(4):1309-1314.
21. Chen CH, *et al.* (2007) A synthetic maternal-effect selfish genetic element drives population replacement in Drosophila. *Science* 316(5824):597-600.
22. Ward CM, *et al.* (2011) Medea selfish genetic elements as tools for altering traits of wild populations: a theoretical analysis. *Evolution* 65(4):1149-1162.
23. Marshall JM & Hay BA (2012) General principles of single construct chromosomal gene drive. *Evolution* 66(7):2150-2166.
24. Burt A (2003) Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc Biol Sci* 270(1518):921-928.
25. Windbichler N, *et al.* (2011) A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* 473(7346):212-215.
26. Gantz VM & Bier E (2015) Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science* 348(6233):442-444.
27. Gantz VM, *et al.* (2015) Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proc Natl Acad Sci U S A* 112(49):E6736-6743.
28. Hammond A, *et al.* (2016) A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nat Biotechnol* 34(1):78-83.
29. Simoni A, *et al.* (2014) Development of synthetic selfish elements based on modular nucleases in Drosophila melanogaster. *Nucleic Acids Res* 42(11):7461-7472.
30. Galizi R, *et al.* (2014) A synthetic sex ratio distortion system for the control of the human malaria mosquito. *Nature communications* 5:3977.
31. Beaghton A, Beaghton PJ, & Burt A (2016) Gene drive through a landscape: Reaction-diffusion models of population suppression and elimination by a sex ratio distorter. *Theoretical population biology* 108:51-69.
32. Marshall JM & Hay BA (2011) Inverse Medea as a novel gene drive system for local population replacement: a theoretical analysis. *J Hered* 102(3):336-341.
33. Marshall JM, Pittman GW, Buchman AB, & Hay BA (2011) Semele: a killer-male, rescue-female system for suppression and replacement of insect disease vector populations. *Genetics* 187(2):535-551.
34. Gould F & Schliekelman P (2004) Population genetics of autocidal control and strain replacement. *Annu Rev Entomol* 49:193-217.
35. Davis S, Bax N, & Grewe P (2001) Engineered underdominance allows efficient and economical introgression of traits into pest populations. *J Theor Biol* 212(1):83-98.
36. Altrock PM, Traulsen A, Reeves RG, & Reed FA (2010) Using underdominance to bi-stably transform local populations. *J Theor Biol* 267(1):62-75.

37. Altrock PM, Traulsen A, & Reed FA (2011) Stability properties of underdominance in finite subdivided populations. *PLoS computational biology* 7(11):e1002260.
38. Akbari OS*, et al.* (2013) A synthetic gene drive system for local, reversible modification and suppression of insect populations. *Current Biol.* 23:671–677.
39. Gokhale CS, Reeves RG, & Reed FA (2014) Dynamics of a combined Medea-underdominant population transformation system. *BMC Evol Biol* 14:98.
40. Reeves RG, Bryk J, Altrock PM, Denton JA, & Reed FA (2014) First steps towards underdominant genetic transformation of insect populations. *PLoS One* 9(5):e97557.
41. Roukos V & Misteli T (2014) The biogenesis of chromosome translocations. *Nat Cell Biol* 16(4):293-300.
42. Coyne JA & Orr HA (2004) *Speciation* (Sinauer Associates, Sunderland, Connecticut).
43. Warburton D (1991) De novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. *Am J Hum Genet* 49(5):995-1013.
44. Ogilvie CM, Braude P, & Scriven PN (2001) Successful pregnancy outcomes after preimplantation genetic diagnosis (PGD) for carriers of chromosome translocations. *Human fertility* 4(3):168-171.
45. Robinson AS (1976) Progress in the use of chromosomal translocations for the control of insect pests. *Biol Rev Camb Philos Soc* 51(1):1-24.
46. Rickards GK (1983) Orientation behavior of chromosome multiples of interchange (reciprocal translocation) heterozygotes. *Annual review of genetics* 17:443-498.
47. Serebrovskii AS (1940) On the possibility of a new method for the control of insect pests. *Zool. Zh.* 19:618-690.
48. Dobzhansky T (1937) *Genetics and the origin of species* (Columbia University Press, NY).
49. Robinson AS & Curtis CF (1973) Controlled crosses and cage experiments with a translocation in Drosophila. *Genetica* 44:591-601.
50. Robinson AS (1976) Progress in the use of chromosomal translocations for the control of insect pests. *Biol. Rev.* 51:1-24.
51. Asman SM, McDonald PT, & Prout T (1981) Field studies of genetic control systems for mosquitoes. *Annu Rev Entomol* 26:289-318.
52. Harewood L & Fraser P (2014) The impact of chromosomal rearrangements on regulation of gene expression. *Hum Mol Genet* 23(R1):R76-82.
53. Harewood L*, et al.* (2010) The effect of translocation-induced nuclear reorganization on gene expression. *Genome Res* 20(5):554-564.
54. Black WCt, Alphey L, & James AA (2011) Why RIDL is not SIT. *Trends Parasitol* 27(8):362-370.
55. Laruson AJ & Reed FA (2016) Stability of underdominant genetic polymorphisms in population networks. *J Theor Biol* 390:156-163.

56. Beumer KJ, Pimpinelli S, & Golic KG (1998) Induced chromosomal exchange directs the segregation of recombinant chromatids in mitosis of Drosophila. *Genetics* 150(1):173-188.
57. Yu Y & Bradley A (2001) Engineering chromosomal rearrangements in mice. *Nat Rev Genet* 2(10):780-790.
58. Egli D, Hafen E, & Schaffner W (2004) An efficient method to generate chromosomal rearrangements by targeted DNA double-strand breaks in Drosophila melanogaster. *Genome Res* 14(7):1382-1393.
59. Forster A*, et al.* (2005) Chromosomal translocation engineering to recapitulate primary events of human cancer. *Cold Spring Harbor symposia on quantitative biology* 70:275-282.
60. Pacher M, Schmidt-Puchta W, & Puchta H (2007) Two unlinked double-strand breaks can induce reciprocal exchanges in plant genomes via homologous recombination and nonhomologous end joining. *Genetics* 175(1):21-29.
61. Torres R*, et al.* (2014) Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. *Nature communications* 5:3964.
62. Jiang J*, et al.* (2016) Induction of site-specific chromosomal translocations in embryonic stem cells by CRISPR/Cas9. *Scientific reports* 6:21918.
63. Chen X, Li M, Feng X, & Guang S (2015) Targeted Chromosomal Translocations and Essential Gene Knockout Using CRISPR/Cas9 Technology in Caenorhabditis elegans. *Genetics* 201(4):1295-1306.
64. Theilmann DA & Stewart S (1992) Molecular analysis of the trans-activating IE-2 gene of Orgyia pseudotsugata multicapsid nuclear polyhedrosis virus. *Virology* 187(1):84-96.
65. Gutierrez E, Wiggins D, Fielding B, & Gould AP (2007) Specialized hepatocyte-like cells regulate Drosophila lipid metabolism. *Nature* 445(7125):275-280.
66. Rong YS & Golic KG (2003) The homologous chromosome is an effective template for the repair of mitotic DNA double-strand breaks in Drosophila. *Genetics* 165(4):1831-1842.
67. Roberts PA (1976) The genetics of chromosome aberration. *The Genetics and Biology of Drosophila*, eds Ashburner M & Novitski E (Academic Press, New York), pp 68-184.
68. Sherizen D, Jang JK, Bhagat R, Kato N, & McKim KS (2005) Meiotic recombination in Drosophila females depends on chromosome continuity between genetically defined boundaries. *Genetics* 169(2):767-781.
69. Deredec A, Burt A, & Godfray HC (2008) The population genetics of using homing endonuclease genes in vector and pest management. *Genetics* 179(4):2013-2026.
70. Unckless RL, Messer PW, Connallon T, & Clark AG (2015) Modeling the Manipulation of Natural Populations by the Mutagenic Chain Reaction. *Genetics* 201(2):425-431.

71. Barton NH & Turelli M (2011) Spatial waves of advance with bistable dynamics: cytoplasmic and genetic analogues of Allee effects. *Am Nat* 178(3):E48-75.
72. Magori K*, et al.* (2009) Skeeter Buster: a stochastic, spatially explicit modeling tool for studying Aedes aegypti population replacement and population suppression strategies. *PLoS Negl Trop Dis* 3(9):e508.
73. Xu C, Legros M, Gould F, & Lloyd AL (2010) Understanding uncertainties in model-based predictions of Aedes aegypti population dynamics. *PLoS Negl Trop Dis* 4(9):e830.
74. Lee Y*, et al.* (2013) Spatiotemporal dynamics of gene flow and hybrid fitness between the M and S forms of the malaria mosquito, Anopheles gambiae. *Proc Natl Acad Sci U S A* 110(49):19854-19859.
75. Fontaine MC*, et al.* (2015) Mosquito genomics. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* 347(6217):1258524.
76. Dickson LB*, et al.* (2016) Reproductive Incompatibility Involving Senegalese Aedes aegypti (L) Is Associated with Chromosome Rearrangements. *PLoS Negl Trop Dis* 10(4):e0004626.
77. Wise de Valdez MR*, et al.* (2011) Genetic elimination of dengue vector mosquitoes. *Proc Natl Acad Sci U S A* 108(12):4772-4775.
78. Deus KM, Saavedra-Rodriguez K, Butters MP, Black WCt, & Foy BD (2012) The effect of ivermectin in seven strains of Aedes aegypti (Diptera: Culicidae) including a genetically diverse laboratory strain and three permethrin resistant strains. *J Med Entomol* 49(2):356-363.
79. Dyck VA, Hendrichs J, & Robinson AS eds (2005) *Sterile Insect Technique: Principles and Practice in Area-Wide Integrated Pest Management* (Springer, Dordrecht, The Netherlands).
80. Harris AF*, et al.* (2012) Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes. *Nat Biotechnol* 30(9):828-830.
81. Carvalho DO*, et al.* (2015) Suppression of a Field Population of Aedes aegypti in Brazil by Sustained Release of Transgenic Male Mosquitoes. *PLoS Negl Trop Dis* 9(7):e0003864.
82. Gibson DG*, et al.* (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* 6(5):343-345.
83. Bischof J, Maeda RK, Hediger M, Karch F, & Basler K (2007) An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci U S A* 104(9):3312-3317.
84. Kyrchanova O, Toshchakov S, Podstreshnaya Y, Parshikov A, & Georgiev P (2008) Functional interaction between the Fab-7 and Fab-8 boundaries and the upstream promoter region in the Drosophila Abd-B gene. *Mol Cell Biol* 28(12):4188-4195.